DEPARTMENT OF STATISTICS
University of Wisconsin
1300 University Ave.
Madison, WI 53706

TECHNICAL REPORT NO. 1122

June 20, 2006

Unified Cortical Surface Morphometry and Its Application to Quantifying Amount of Gray Matter

Moo K. Chung

mchung@stat.wisc.edu
http://www.stat.wisc.edu/∼mchung

Department of Statistics
Department of Biostatistics and Medical Informatics
Keck Laboratory for Functional Brain Imaging and Behavior
University of Wisconsin-Madison
Madison, WI 53706

Unified Cortical Surface Morphometry and Its Application to Quantifying Amount of Gray Matter

Moo K. Chung, Kim M. Dalton, Li Shen,
Alan C. Evans, Richard J. Davidson

June 20, 2006

Abstract

In quantifying the amount of gray matter in a population, voxel-based morphometry (VBM) and cortical thickness analysis are the two most widely used techniques. There are still many unanswered methodological questions regarding these methods. It is not clear what is the optimal amount of registration needed in VBM. It is assumed that the gray matter density obtained in the VBM will be positively correlated to cortical thickness but there is no systematic study that compared these two disparate measurements. The gray matter density is a 3D measurement defined in voxels in 3D whole brain volume while the thickenss is a 1D measurement defined along the 2D cortical surface. Due to this disparate dimensionality, it is not clear how to compare them in a systematic fashion.

In this paper, we present a novel weighted Fourier Series (WFS) representation for the cortical surface that enables us to address these questions in a unified mathematical framework. The WFS representation is a data smoothing technique formulated as a solution of the Cauchy problem. This representation provides the explicit smooth functional estimation of true unknown cortical boundary as a linear combination of basis functions. Based on this new representation, cortical thickness and gray matter density can be defined and compared naturally. The basic theory underlying the weighted Fourier Series representation and its numerical implementation issues are presented in detail.

As an illustration, this unified approach is applied in the cortical thickness analysis and the VBM in a group of autistic subjects. The thickness analysis and VBM results are compared in parallel.

1 Introduction

The cerebral cortex has a highly convoluted geometry and it is likely that the local difference in gray matter concentration can characterize a clinical population. Among many magnetic resonance imaging (MRI) morphometric techniques proposed in quantifying the amount of gray matter locally, voxel-based morphometry (VBM) [2] [20] [57] and cortical thickness analysis [11] [13] [23] [40] have been the two most widely used techniques so far.

Voxel-based morphometry (VBM) is a fully automated image analysis technique allowing identification of regional differences in gray and white matters between groups of subjects without a prior region of interest (ROI). Although there are many variations in VBM, the underpinning procedures for different VBM are identical. The 3D whole brain MRI is normalized into a template then each voxel is assigned a probability of that voxel belonging to a particular tissue class. There are three main tissue classes: cerebrospinal fluid (CSF), gray matter, and white matter. Afterwards the tissue probability is used in a general linear model (GLM) [26] to characterizing the amount of tissue concentration at each voxel.

On the other hand, cortical thickness analysis requires an additional step of segmenting the cortical surface. The CSF/gray matter interface is called the *outer surface* (pial surface) while the gray/white matter interface is called the *inner surface* [36]. Then the distance between the outer and inner surfaces is defined as the *cortical thickness*. The cortical thickness has been widely used as an anatomical index for quantifying the amount of gray matter in the brain [11] [13] [23].

Although VBM and cortical thickness analysis are popular techniques, in almost all studies, only a single morphometric technique is used to quantify gray matter. Consequently, each study provides only a small vignette of the overall picture of a population. Further, studies using different morphometric techniques often result in contradictory findings. Therefore it is necessary to be able to compare and the VBM to the cortical thickness analysis directly in a unified methodological framework. Studies that effectively compare different morphometric techniques and anatomical measures will provide more insight into understanding the population.

As a basic mathematical tool for achieving these goals, we present a new representation technique called the weighed Fourier series (WFS) representation. This is an explicit data smoothing technique formulated as a solution to a partial differential equation (PDE) and related to the spherical harmonic (SPHARM) representation [27] [49] indirectly. The basic theoretical properties of WFS and its numerical implementation issues will be

presented in great detail so that readers should be able to implement and possibly modify the WFS for their own use. Based on WFS, VBM and cortical thickness analysis will be performed in parallel and the results will be compared and combined together.

The two major contributions of the paper is the theoretical and computational development of WFS, and showing various applications toward VBM and cortical thickness analysis in a unified framework.

2 Literature Review

We will mainly review literatures that are directly related to our methodology and address what is our new contributions in the context of the previous literatures. Although there are many different morphometric techniques such as deformation-based morphometry (DBM) [13] [21] and tensor-based morphometry (TBM) [4] [12] [13] [51] [52], we will not review or address these methods.

2.1 Voxel-based Morphometry

VBM as implemented in the statistical parametric mapping (SPM) software (http://www.fil.ion.ucl.ac.uk/spm) starts with normalizing each structural MRI to the standard SPM template and segmenting it into white and gray matters, and CSF based on a Gaussian mixture model [2] [20] [29] [57]. In a slightly different formulation, the tissue density is generated by convoluting the binary mask of the tissue with a 3D Gaussian kernel [42]. The resulting density maps are warped into a normalized space and compared across subjects. A modified version of VBM has been also performed along the cortex, where a fraction of gray matter within a ball of radius 15mm is taken as gray matter density [51]. This is equivalent to convoluting the binary mask of the gray matter with a uniform probability distribution of radius 15mm and interpolating voxel values to the cortical surface mesh. This equivalence relation is the basis of how we will project the 3D density maps to the 2D cortical surface and compare them with the cortical thickness in our study.

VBM has been applied to various anatomical studies: normal development [29] [42], autism [10], depression [43], epilepsy [37] and Alzheimer's disease [32] [51]. Most previous studies in VBM did not compare the tissue density to cortical thickness or any other morphometric measures so our study is the first to present the comparative analysis with some counterintuitive results.

The objective of VBM is to compare regional difference in relative tissue concentration. It is not necessary for image registration used in VBM to match every cortical features exactly, but merely corrects for global brain shape differences. If the image registration was exact, all the segmented images would appear identical and no statistically significant differences would be detected [2]. The amount of image registration needed in VBM has been a contentious issue that has yet to be addressed quantitatively [3] [5] [16] [38]. The WFS representation will enable us to address this issue in a unified fashion.

2.2 Cortical Thickness Analysis

The cerebral cortex has the topology of a 2-dimensional convoluted sheet. Most of the features that distinguish these cortical regions can only be measured relative to that local orientation of the cortical surface [17]. Unlike 3D whole brain volume based VBM, 1D cortical thickness measures have the advantage of providing a direct quantification of cortical geometry. It is likely that different clinical populations will exhibit different cortical thickness. By analyzing cortical thickness, brain shape differences can be quantified locally [13] [24] [36] [40].

The cortical surfaces are usually segmented as triangle meshes that are constructed from deformable surface algorithms [19] [23] [36]. Then the cortical thickness is mainly defined and estimated as the shortest distance between vertices of the two triangle meshes [23] [36]. The mesh construction and discrete thickness computation procedures introduce substantial noise in the thickness measure [11] (Figure 5). So it is necessary to increase the signal-to-noise ratio (SNR) and smoothness of data for the subsequent random field based statistical analysis. For smoothing cortical data, diffusion equation based methods have been used [1] [8] [11] [13]. The shortcoming of these approaches is the need for numerically solving the diffusion equation possibly via the finite element technique. This is an additional time computational step on top of the cortical thickness estimation. In this paper, we present a more direct approach that smoothes and parameterizes the coordinates of a mesh directly via WFS such that the resulting thickness measures are already smooth. In the WFS, the cortical surfaces are estimated as a weighted linear combination of smooth basis functions so that most algebraic operations on the WFS will also be smooth. This is the first study comparing the result of VBM to a cortical thickness analysis. Taking the WFS as some sort of ground truth, gray matter density and cortical thickness are correlated both within a subject and between subjects.

2.3 Spherical Harmonic Representation

The SPHARM representation [6] has been applied to subcortical structures such as the hippocampus and the amygdala [27] [31] [34] [49]. In particular, Gerig et al. used the mean squared distance (MSD) of SPHARM coefficients in quantifying ventricle surface shape in a twin study [27]. Shen et al. used the principal component analysis technique on the SPHARM coefficients of schizophrenic hippocampal surfaces in reducing the data dimension [49]. Recently it has begun to be applied to more complex cortical surfaces [12] [31] [48]. Gu et al. presented the SPHARM representation as a surface compression technique, where the main geometric feasures are encoded in the low degree spherical harmonics, while the noises are in the high degree spherical harmonics [31].

In SPHARM, the spherical harmonic functions are used in constructing the Fourier series expansion of the mapping from the cortex to a unit sphere. So the SPHARM representation is more of an interpolation technique than a smoothing technique. On the other hand, WFS is a kernel smoothing technique given as a solution to a particular PDE. The solution to the PDE is expanded in basis functions similar to the Fourier series expansion with weights. WFS offers many advantages over previous PDE-based smoothing techniques. The PDE-based smoothing methods tend to suffer a numerical convergence problem [1] [13] while WFS has no such problem. Since the traditional PDE-based smoothing gives an implicit numerical solution, setting up a statistical model is not straightforward. However, WFS provides an explicit series expansion so it is easy to apply wide variety of statistical modeling techniques such as the GLM [26] , principal component analysis (PCA) [49] and functional-PCA [41] [46]. Bulow used the spherical harmonics in isotropic heat diffusion via the Fourier transform on a unit sphere as a form of hierarchical surface representation [7].

The SPHARM representation will be shown to be the special case of WFS. In SPHARM, all measurements are assigned equal weights and the coefficients of the series expansion is estimated in the least squares fashion. In WFS, closer measurements are weighted more and the coefficients of the series expansion is estimated in the weighted least squares fashion. So WFS is more suitable than SPHARM when the realization of the cortical boundaries, as triangle meshes, are noisy [12]. In most SPHARM literatures, the degree of the Fourier series expansion has been arbitrary determined and the problem of the optimal degree has not been addressed. Our WFS formulation addresses the determination of the optimal degree in a unified statistical modeling framework. The WFS-based global parametrization is computa-

tionally expensive compared to the local quadratic polynomial fitting [6] [13] [17] [33] [45] while providing more accuracy and flexibility for hierarchical representation.

3 Cauchy problem as a smoothing process

Consider $\mathcal{M} \in \mathbb{R}^d$ to be a compact differentiable manifold. Let $L^2(\mathcal{M})$ be the space of square integrable functions in $\mathcal{M}$ with inner product

$$\langle g_1, g_2 \rangle = \int_{\mathcal{M}} g_1(p) g_2(p) \, d\mu(p), \tag{1}$$

where μ is the Lebegue measure such that $\mu(\mathcal{M})$ is the total volume of $\mathcal{M}$. The norm $\| \cdot \|$ is defined as

$$\|g\| = \langle g, g \rangle^{1/2}.$$

The linear partial differential operator $\mathcal{L}$ is *self-adjoint* if

$$\langle g_1, \mathcal{L} g_2 \rangle = \langle \mathcal{L} g_1, g_2 \rangle$$

for all $g_1, g_2 \in L^2(\mathcal{M})$. Then the eigenvalues λ_j and eigenfunctions ψ_j of the operator $\mathcal{L}$ are obtained by solving

$$\mathcal{L}\psi_j = \lambda_j \psi_j. \tag{2}$$

Without the loss of generality, we can order eigenvalues

$$0 < \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \cdots$$

and the eigenfunctions to be orthonormal with respect to the inner product (1). Consider the Cauchy problem of the following form:

$$\partial_t g + \mathcal{L} g = 0, g(p, t = 0) = f(p). \tag{3}$$

The initial functional data $f(p)$ can be further stochastically modeled as

$$f(p) = \nu(p) + \epsilon(p), \tag{4}$$

where ϵ is a stochastic noise modeled as a mean zero Gaussian random field and ν is the unknown signal to be estiamted. The PDE (3) diffuses noisy initial data f over time and estimate the unknown signal ν as a solution. The time t controls the amount of smoothing and will be termed as the *bandwidth*. The unique solution to equation (3) is given by the following theorem.

Theorem 1 *For the self-adjoint linear differential operator $\mathcal{L}$, the unique solution of the Cauchy problem (3) is given by*

$$g(p,t) = \sum_{j=0}^{\infty} e^{-\lambda_j t} \langle f, \psi_j \rangle \psi_j(p). \tag{5}$$

Proof For each fixed t, g has expansion

$$g(p,t) = \sum_{j=0}^{\infty} c_j(t) \psi_j(p). \tag{6}$$

Substitute equation (6) into (3). Then we obtain

$$\partial_t c_j(t) + \lambda_j c_j(t) = 0. \tag{7}$$

The solution of equation (7) is given by $c_j(t) = b_j e^{-\lambda_j t}$. So we have solution

$$g(p,t) = \sum_{j=0}^{\infty} b_j e^{-\lambda_j t} \psi_j(p).$$

At $t = 0$, we have

$$g(p,0) = \sum_{j=0}^{\infty} b_j \psi_j(p) = f(p).$$

The coefficients b_j must be the Fourier coefficients $\langle f, \psi_j \rangle$.

The implication of Theorem 1 is obvious. The solution decreases exponentially as time t increases and smoothes out high spatial frequency noises much faster than low frequency noises. This is the basis of many of PDE-based image smoothing methods. PDE involving self-adjoint linear partial differential operators such as the Laplace-Beltrami operator or iterated Laplacian have been widely used in medical image analysis as a way to smooth either scalar or vector data along anatomical boundaries [1] [7] [8] [13]. These methods directly solve the PDE using standard numerical techniques such as the finite difference method or the finite element method. The problem with directly solving PDEs is the numerical instability and the complexity of setting up the numerical scheme. WFS differs from these previous method in such a way that we only need to estimate the Fourier coefficients in a hierarchical fashion to solve the PDE.

3.1 Weighted Fourier Series

We will investigate the properties of the finite expansion of (5) denoted by

$$\mathcal{F}_t^k[f](p) = \sum_{j=0}^{k} e^{-\lambda_j t} \langle f, \psi_j \rangle \psi_j(p).$$

This expansion will be called as the *weighted Fourier Series* (WFS). By re-arranging the inner product, the WFS can be rewritten as kernel smoothing:

$$\begin{aligned} \mathcal{F}_t^k[f](p) &= \sum_{j=0}^{k} e^{-\lambda_j t} \psi_j(p) \int_{S^2} f(q) \psi_j(q) \, d\mu(q) && (8) \\ &= \int_{S^2} f(q) K_t^k(p,q) \, d\mu(q) && (9) \end{aligned}$$

with symmetric positive definite kernel K_t^k given by

$$K_t^k(p,q) = \sum_{j=0}^{k} e^{-\lambda_j t} \psi_j(p) \psi_j(q). \tag{10}$$

The subscript t is introduced to show the dependence of the kernel on time t. This shows that the solution of the Cauchy problem (3) can be interpreted as kernel smoothing.

When the differential operator $\mathcal{L} = \Delta$, the Laplace-Beltrami operator, the Cauchy problem (3) becomes an isotropic diffusion equation. For this particular case, K_t^∞ is called the *heat kernel* with bandwidth t [9] [11]. For an arbitrary cortical manifold, the basis functions ψ_j can be computed and the exact shape of heat kernel can be determined numerically. Although it can be done by setting up a huge finite element method [44], this is not a trivial numerical computation. A simpler approach is to use the first order approximation of the heat kernel for small bandwidth and iteratively apply it up to the desired bandwidth [11].

The WFS can be reformulated as a kernel regression problem [22]. At each fixed point p, we estimate unknown signal ν (4) with smooth function $h \in L^2(\mathcal{M})$ by minimizing the integral of the weighted squared distance between f and h:

$$\min_{h \in L^2(\mathcal{M})} \int_{\mathcal{M}} K_t(p,q) \big| f(q) - h(p) \big|^2 \, d\mu(q). \tag{11}$$

The minimizer of (11) is given by the following theorem.

Theorem 2

$$\begin{aligned} & \frac{\mathcal{F}_t^k[f](p)}{\int_{\mathcal{M}} K_t^k(p,q)\, d\mu(q)} \\ = \quad & \arg\min_{h\in L^2(\mathcal{M})} \int_{\mathcal{M}} K_t^k(p,q)\big|f(q)-h(p)\big|^2\, d\mu(q). \end{aligned}$$

Proof Since the integral is quadratic in h, the minimum exists and obtained when

$$\begin{aligned} & \frac{\partial}{\partial h}\int_{\mathcal{M}} K_t^k(p,q)|f(q)-h(p)|^2\, d\mu(q) \\ = \quad & -2\int_{\mathcal{M}} K_t^k(p,q)[f(q)-h(p)]\, d\mu(q) = 0. \end{aligned}$$

Solving the equation, we obtain the result.

Theorem 2 shows the WFS is the solution of a weighted least squares minimization problem.

When $\mathcal{L}$ is the Laplace-Beltrami operator with $k=\infty$, the heat kernel K_t^∞ is a probability distribution in $\mathcal{M}$, i.e.

$$\int_{\mathcal{M}} K_t^\infty(p,q)\, d\mu(p) = 1.$$

For this special case, Theorem 2 simplifies to

$$\mathcal{F}_t^\infty[f](p) = \arg\min_{h\in L^2(\mathcal{M})} \int_{\mathcal{M}} K_t^\infty(p,q)\big|f(q)-h(p)\big|^2\, d\mu(q).$$

In minimizing the weighted least squares in Theorem 2, it is possible to restrict the function space $L^2(\mathcal{M})$ to a finite subspace that is more useful in numerical implementation. Let

$$\mathcal{H}_l = \{\sum_{j=0}^{l} \beta_j\psi_j(p) : \beta_j \in \mathbb{R}\}$$

be the subspace spanned by basis $\psi_0, \cdots, \psi_l$. Then we have the following theorem.

Theorem 3 *If* $\int_{\mathcal{M}} K_t^k(p,q)\, d\mu(q) = 1$ *and* $l \le k$, *then*

$$\mathcal{F}_t^l[f](p) = \arg\min_{h\in\mathcal{H}_l} \int_{\mathcal{M}} K_t^k(p,q)|f(q)-h(p)|^2\, d\mu(q).$$

Proof Let $h(p) = \sum_{j=0}^{l} \beta_j \psi_j(p) \in \mathcal{H}_l$. The integral is written as

$$I(\beta_0, \cdots, \beta_l) = \int_{\mathcal{M}} K_t^k(p,q) \Big| f(q) - \sum_{j=0}^{l} \beta_j \psi_j(p) \Big|^2 \, d\mu(q).$$

Since the functional I is quadratic in $\beta_0, \cdots, \beta_l$, the minimum exists and it is obtained when $\frac{\partial I}{\partial \beta_{j'}} = 0$ for all j'. By differentiating I and rearranging terms, we obtain

$$\begin{aligned} & \sum_{j=0}^{k} e^{-\lambda_j t} \psi_j(p) \psi_{j'}(p) \int_{\mathcal{M}} f(q) \psi_j(q) \, d\mu(q) \\ = & \sum_{j=0}^{l} \beta_j \psi_j(p) \psi_{j'}(p). \end{aligned}$$

Now integrate the equations respect to measure $\mu(p)$ and obtain

$$e^{-\lambda_{j'} t} \langle f, \psi_{j'} \rangle = \beta_{j'}.$$

If K_t^k is a probability distribution, this theorem holds. For any other symmetric positive definite kernel, it can be made to be a probability distribution by renormalizing it. So Theorem 3 can be applicable in wide variety of kernels.

3.2 Isotropic Diffusion on Unit Sphere

Let us apply the WFS theory to a unit sphere. Since algebraic surfaces provide basis functions in a close form, it is not necessary to construct numerical basis [44]. The WFS in S^2 is given by the solution of isotropic diffusion. The spherical parametrization of S^2 is given by the polar angle θ and the azimuthal angel φ:

$$p = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta) \tag{12}$$

with $p = (\theta, \varphi) \in [0, \pi] \otimes [0, 2\pi)$. The spherical Laplacian Δ corresponding to the parametrization (12) is given by

$$\Delta = \frac{1}{\sin\theta} \frac{\partial}{\partial\theta} \Big(\sin\theta \frac{\partial}{\partial\theta} \Big) + \frac{1}{\sin^2\theta} \frac{\partial^2}{\partial^2 \varphi}.$$

There are $2l+1$ eigenfunctions $Y_{lm} (-l \le m \le l)$, corresponding to the same eigenvalue $\lambda_l = l(l+1)$ satisfying

$$\Delta Y_{lm} = \lambda_l Y_{lm}.$$

Y_{lm} is called the *spherical harmonic* of degree l and order m [15, 53]. It is given explicitly as

$$Y_{lm} = \begin{cases} c_{lm} P_l^{|m|}(\cos\theta)\sin(|m|\varphi), & -l \le m \le -1, \\ \frac{c_{lm}}{\sqrt{2}} P_l^0(\cos\theta), & m = 0, \\ c_{lm} P_l^{|m|}(\cos\theta)\cos(|m|\varphi), & 1 \le m \le l, \end{cases}$$

where $c_{lm} = \sqrt{\frac{2l+1}{2\pi}\frac{(l-|m|)!}{(l+|m|)!}}$ and P_l^m is the associated Legendre polynomials of order m.

Unlike many previous imaging literatures on spherical harmonics that used the complex-valued spherical harmonics [7] [27] [31] [49], only real-valued spherical harmonics with different normalizing constants are used throughout the paper for its convenience for setting up a real-valued stochastic model.

For $f, h \in L^2(S^2)$, we define the inner product as

$$\langle f, h \rangle = \int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\pi} f(p)h(p)\, d\mu(p),$$

where Lebesgue measure $d\mu(p) = \sin\theta d\theta d\varphi$. Then with respect to the inner product, the spherical harmonics satisfies the orthonormal condition

$$\int_{S^2} Y_{ij}(p) Y_{lm}(p)\, d\mu(p) = \delta_{il}\delta_{jm},$$

where δ_{il} is the Kroneker's delta. The kernel K_t^k is given by

$$K_t^k(p, q) = \sum_{l=0}^{k} \sum_{m=-l}^{l} e^{-l(l+1)t} Y_{lm}(p) Y_{lm}(q). \tag{13}$$

The associated WFS is given by

$$\mathcal{F}_t^k[f](p) = \sum_{l=0}^{k} \sum_{m=-l}^{l} e^{-l(l+1)t} f_{lm} Y_{lm}(p)$$

with Fourier coefficient $f_{lm} = \langle f, Y_{lm} \rangle$. This form of WFS is called the *weighted-SPHARM* and it has been used as a global differential parameterization of the cortex for tensor-based morphometry [12]. The special case $\mathcal{F}_0^k[f]$ is the traditional SPHARM representation used in representing the

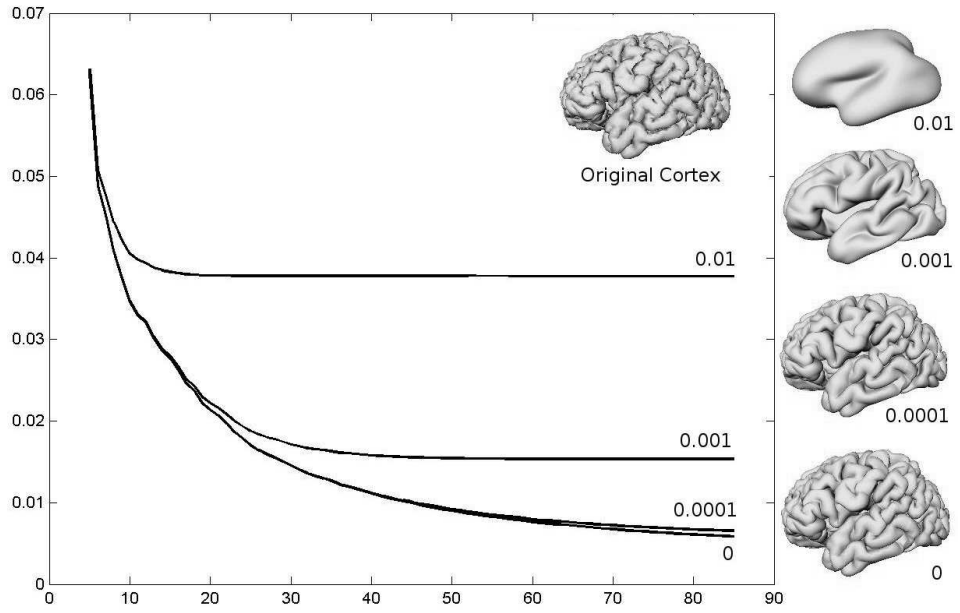


Figure 1: Plots of RMSE over degree for bandwidths $t = 0.01$, 0.001, 0.0001, 0. The smoothed outer surfaces are at degree $k = 85$ showing the bandwidth controls the amount of smoothing in representing the cortex. The bandwidth $t = 0$ corresponds to the traditional SPHARM. As $t \to 0$, the WFS converges to SPHARM.

Cartesian coordinates of anatomical boundaries [27] [31] [49]. Consider subspace

$$\mathcal{H}_k = \{\sum_{l=0}^{k} \sum_{m=-l}^{l} \beta_{lm} Y_{lm} : \beta_{lm} \in \mathbb{R}\} \subset L^2(S^2),$$

which is spanned by up to the k-th degree spherical harmonics. Then the SPHARM satifies the least squares minimization problem different from Theorem 2 and Theorem 3.

Theorem 4

$$\mathcal{F}_0^k[f](p) = \arg\min_{h \in \mathcal{H}_k} \|f - h\|^2. \tag{14}$$

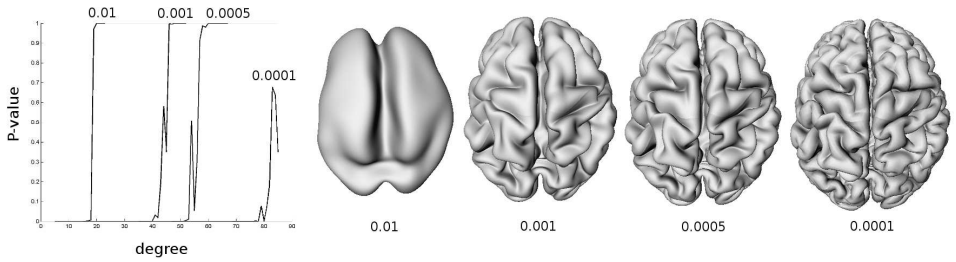


Figure 2: Automatic degree selection in the IRF algorithm. For each bandwidth t, the optimal degree is automatically selected by checking if adding an additional degree will be statistically significant. The outer cortical surfaces are the results of the optimal selection procedure. The optimal degrees are $k = 18(t = 0.01), k = 42(t = 0.001), k = 52(t = 0.0005), k = 78(t = 0.0001)$. For our study, $t = 0.0001$ and the corresponding degree $k = 78$ is used through the paper.

4 Numerical Implementation

In constructing the WFS representation, all we need is estimating Fourier coefficients $f_{lm} = \langle f, Y_{lm} \rangle$. There are three major techniques for computing the Fourier coefficients. The first method numerically integrate the Fourier coefficients over a high resolution triangle mesh [9]. Although this approach is the simplest to implement numerically and more accurate, due to its brute force nature of the technique, the computation is extremely slow. The second method is based on the fast Fourier transform (FFT) [7] [31]. The drawback of the FFT is the need for a predefined regular grid system so if the mesh topology is different for each subject as in the case of FreeSurfer [23], a time consuming interpolation is needed. The third method is based on solving a system of linear equations [27] [48] [49] that minimize the least squares problem in Theorem 4. This is the most widely used numerical technique in SPHARM literatures. However, the direct application of the least squares estimation is not desirable when the size of the linear equation is extremely large.

Let

$$h = \sum_{l=0}^{k} \sum_{m=-l}^{l} \beta_{lm} Y_{lm}(p) \in \mathcal{H}_k.$$

Given n nodes $p_1, \cdots, p_n$ in S^2 mesh, the discretization of (14) is given by

$$\|f - h\|^2 \approx \sum_{i=1}^{n} \Big[f(p_i) - \sum_{l=0}^{k} \sum_{m=-l}^{l} \beta_{lm} Y_{lm}(p_i) \Big]^2. \tag{15}$$

The minimum of (15) is obtained when

$$f(p_i) = \sum_{l=0}^{k} \sum_{m=-l}^{l} \beta_{lm} Y_{lm}(p_i) \tag{16}$$

all $1 \leq i \leq n$. The equation (16) is referred as the *normal equation* in statistical literatures. The normal equation is usually solved via a matrix inversion. Let

$$\boldsymbol{f} = (f(p_1), \cdots, f(p_n))'$$

and

$$\boldsymbol{\beta}_l = (\beta_{l,-l}, \cdots, \beta_{l,l})'.$$

Also let

$$\mathbf{Y}_l = \begin{bmatrix} Y_{l,-l}(p_1) & \cdots & Y_{l,l}(p_1) \\ \vdots & \ddots & \vdots \\ Y_{l,-l}(p_n) & \cdots & Y_{l,l}(p_n) \end{bmatrix}$$

be a $n \times (2l+1)$ submatrix consisting of the l-th degree spherical harmonics evaluated at each node p_i. Then (16) can be rewritten in the following matrix form:

$$\boldsymbol{f} = \mathbf{Y}\boldsymbol{\beta}. \tag{17}$$

with the *design matrix* $\mathbf{Y} = [\mathbf{Y}_0, \mathbf{Y}_1, \cdots, \mathbf{Y}_k]$ and unknown parameter vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_0', \cdots, \boldsymbol{\beta}_k')'$. The linear system is solved via

$$\boldsymbol{\beta} = (\mathbf{Y}'\mathbf{Y})^{-1}\mathbf{Y}'\boldsymbol{f}. \tag{18}$$

The problem with this widely used formulation is that the size of the matrix $\mathbf{Y}$ is $n \times (k+1)^2$, which becomes fairly large and may not fit in most of computer memories. So it becomes unpractical to perform matrix operation (18) directly. This is true for many cortical surface extraction tools such as FreeSurfer [23] that produces no less than $n > 100,000$ nodes for each hemisphere. This computational bottleneck can be overcome by breaking the least squares problem in the subspace $\mathcal{H}_k$ into smaller subspaces using the *iterative residual fitting* (IRF) algorithm [48]. The IRF for WFS will

be given in the following subsection. Although IRF was first introduced in [48], the correctness of the algorithm was not given in [48]. In this paper, we present Theorem 5 that proves the correctness of the IRF for the first time. The IRF can be also used in estimating SPHARM coefficients by letting the bandwidth $t = 0$ in the algorithm.

4.1 Iterative residual fitting (IRF) algorithm

Decompose the subspace $\mathcal{H}_k$ into smaller subspaces as the direct sum:

$$\mathcal{H}_k = \mathcal{I}_0 \oplus \mathcal{I}_1 \cdots \oplus \mathcal{I}_k,$$

where subspace

$$\mathcal{I}_l = \{ \sum_{m=-l}^{l} \beta_{lm} Y_{lm}(p) : \beta_{lm} \in \mathbb{R} \}$$

is spanned by the l-th degree spherical harmonics only. Then the IRF estimates the Fourier coefficients $\boldsymbol{\beta}_l$ in each subspace $\mathcal{I}_l$ iteratively from increasing the degree from 0 to k. Suppose we estimated the coefficients $\boldsymbol{\beta}_0, \cdots, \boldsymbol{\beta}_{l-1}$ up to degree $l-1$ somehow. Then the residual vector $\boldsymbol{r}_{l-1}$ based on this estimation is given by

$$\boldsymbol{r}_{l-1} = \boldsymbol{f} - \sum_{j=0}^{l-1} e^{-j(j+1)t} \mathbf{Y}_j \boldsymbol{\beta}_j. \tag{19}$$

The components of the residual vector $\boldsymbol{r}_{l-1}$ are identical so we denote all of them as r_{l-1}. At the next degree l, we estimate the coefficients $\boldsymbol{\beta}_l$ by minimizing the difference between the residual r_{l-1} and $\sum_{m=-l}^{l} \beta_{lm} Y_{lm} \in \mathcal{I}_l$. This is formally stated as the following theorem:

Theorem 5

$$\sum_{m=-l}^{l} f_{lm} Y_{lm} = \arg\min_{h \in \mathcal{I}_l} \|r_{l-1} - h\|^2. \tag{20}$$

Proof Let $h = \sum_{m=-l}^{l} \beta_{lm} Y_{lm}(p) \in \mathcal{I}_l$. The squared norm is denoted as

$$I(\beta_{l,-l}, \cdots, \beta_{l,l}) = \int_{S^2} \Big[\boldsymbol{r}_{l-1}(p) - \sum_{m=-l}^{l} \beta_{lm} Y_{lm}(p) \Big]^2 \, d\mu(p).$$

Since I is quadratic in β_{lm}'s, the minimum is obtained when $\frac{\partial I}{\partial \beta_{lm'}} = 0$ for all m'. By differentiating I with respect to $\beta_{lm'}$ and letting it equal to zero, we obtain

$$\begin{aligned} & \int_{S^2} Y_{lm'}(p) \Big[\boldsymbol{r}_{l-1}(p) - \sum_{m=-l}^{l} \beta_{lm} Y_{lm}(p) \Big] \, d\mu(p) \\ = & \int_{S^2} Y_{lm'}(p) \boldsymbol{r}_{l-1}(p) \, d\mu(p) - \beta_{lm'}. \end{aligned} \tag{21}$$

From (19), we have

$$r_{l-1} = f - \sum_{j=0}^{l-1} \sum_{m=-j}^{j} e^{-j(j+1)t} \beta_{jm} Y_{jm},$$

which is a linear combination of spherical harmonics up to $(l-1)$-th degree so it is orthonormal to Y_{lm}. Then the first term in (21) simplifies to

$$\int_{S^2} Y_{lm'}(p) \boldsymbol{r}_{l-1}(p) \, d\mu(p) = \langle f, Y_{lm'} \rangle.$$

Theorem 5 proves that the correctness of IRF procedure. Then the discretization and the optimization is based on the normal equation approach (14):

$$\boldsymbol{\beta}_l = (\mathbf{Y}_l \mathbf{Y}_l)^{-1} \mathbf{Y}_l' \boldsymbol{r}_{l-1}.$$

Summarizing the results, the IRF algorithm is given below.

Algorithm 1 *Iterative Residual Fitting (IRF)*

1. *Let* $l = 0$.
2. $\boldsymbol{\beta}_0 \leftarrow (\mathbf{Y}_0 \mathbf{Y}_0)^{-1} \mathbf{Y}_0' \boldsymbol{f}$.
3. $l \leftarrow l + 1$.
4. $\boldsymbol{r}_{l-1} \leftarrow \boldsymbol{f} - \sum_{j=0}^{l-1} e^{-j(j+1)t} \mathbf{Y}_j \boldsymbol{\beta}_j$.
5. $\boldsymbol{\beta}_l \leftarrow (\mathbf{Y}_l \mathbf{Y}_l)^{-1} \mathbf{Y}_l' \boldsymbol{r}_{l-1}$.
6. *If* $l < k$, *go to step 3.*

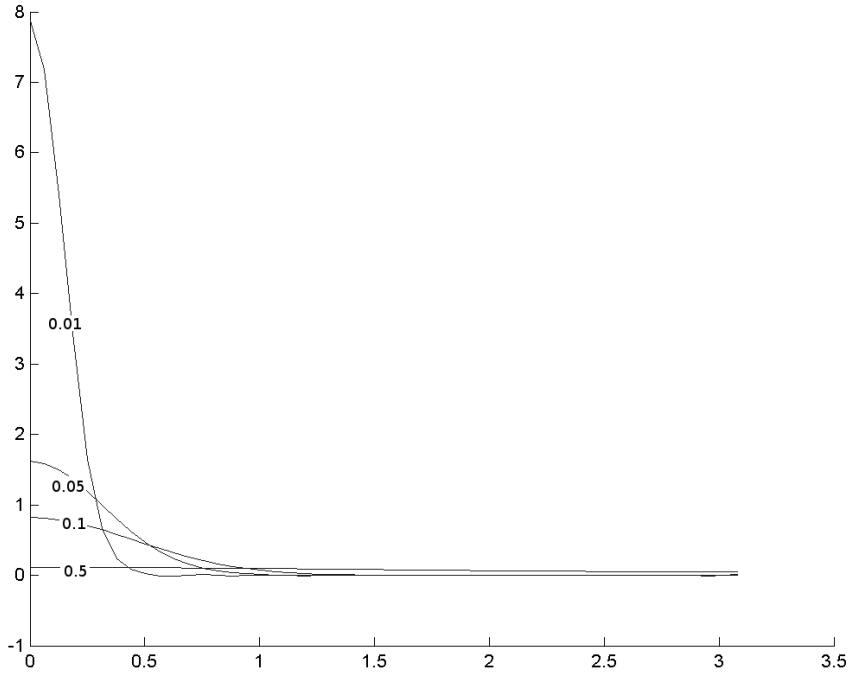


Figure 3: Shape of heat kernel with different bandwidth $t = 0.01$, 0.05, 0.1, 0.5. The horizontal axis is the θ value from the north pole ($\theta = 0$) to the south pole ($\theta = \pi$). The weighting scheme used in WFS follows the shape of the heat kernel.

4.2 Automatic optimal degree selection in IRF

The IRF algorithm hieratically build the WFS from lower to higher degree. In most previous SPHARM literatures [7] [28] [27] [31] [48] [49], the issue of the optimal degree has not been addressed. We present a statistical framework for automatically determining optimal degree in the Algorithm 1. Although increasing the degree of WFS increases the goodness-of-fit, it also increases the number of coefficients to be estimated quadratically. So it is necessary to find the optimal degree where the goodness-of-fit and the number of parameters balance out.

The Fourier coefficients f_{lm} can be modeled to follow independent normal distribution $N(\mu_{lm}, \sigma_l^2)$. It is natural to assume the equal variance within the same degree. This assumption is equivalent to modeling WFS as the

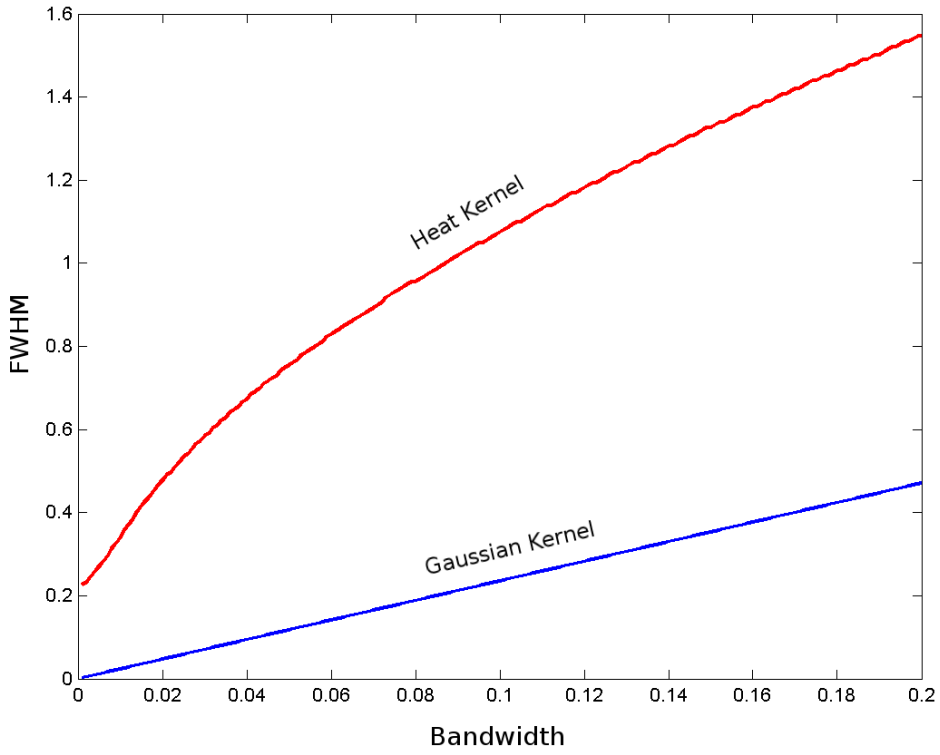


Figure 4: Plot of FWHM (vertical) over bandwidth t (horizontal) for both heat kernel and Gaussian kernel. The FWHM has to be numerically estimated in the case of the heat kernel. The numerically computed FWHM is used in the random field theory based multiple comparison correction.

sum of signal plus noise:

$$f(p_i) = \sum_{l=0}^{k} \sum_{m=-l}^{l} e^{-\lambda(\lambda+1)t} \mu_{lm} Y_{lm}(p_i) + \boldsymbol{\epsilon}(p_i), \tag{22}$$

where $\boldsymbol{\epsilon}$ is a zero mean isotropic Gaussian random field.

Then at each iteration, we test if adding the k-th degree terms in the $(k-1)$-th degree model is statistically significant by testing the null hypothesis

$$H_0 : \mu_{km} = 0 \text{ for } |m| \leq k.$$

The test statistic is constructed from the sum of squared errors (SSE). Let

the k-th degree *sum of squared errors* (SSE) be

$$\mathrm{SSE}_k = \sum_{i=1}^{n} \boldsymbol{r}_k^2(p_i).$$

The plot of the *root mean squared errors* (RMSE), $\sqrt{\mathrm{SSE}_k/n}$ for varying degree $5 \leq k \leq 85$ is shown in Figure 1. As the degree k increases, the root mean squared errors keep decreasing until it flattens out. So it is reasonable to stop the iteration when the decrease in error is no longer significant. Under H_0, the test statistic is

$$F = \frac{(\mathrm{SSE}_{k-1} - \mathrm{SSE}_k)/(2k+1)}{\mathrm{SSE}_{k-1}/(n-(k+1)^2)} \sim F_{2k+1,n-(k+1)^2},$$

the F-distribution with $2k+1$ and $n-(k+1)^2$ degrees of freedom. We compute the F statistic at each degree and stop the IRF procedure if the corresponding P-value first becomes bigger than the pre-specified significance α ($\alpha = 0.01$ in this study). For bandwidth $t = 0.0001$, the optimal degree is determined to be $k = 78$ (Figure 2).

4.3 Computing FWHM

Since the WFS representation is a kernel smoothing method, it is useful to know the full width at the half maximum (FWHM) of the underlying kernel. The computed FWHM is later used in the random field based multiple comparison corrections [11] [13] [56] [55]. Computing the FWHM of heat kernel used in WFS is not trivial since there is no known close form expression for FWHM as a function of bandwidth t. Therefore, FWHM is computed numerically.

For $p, q, r \in S^2$, let us define the Cartesian inner product $\cdot$ as $p \cdot q = \cos(\theta)$, where θ is an angle between p and q. The heat kernel (13) is symmetric along the geodesic circle. If $p \cdot q = p \cdot r$, we have $K_t^k(p,q) = K_t^k(p,r)$. This property can be used to simplify the expansion (13) using the harmonic addition theorem [30] [53].

Theorem 6 *(Harmonic addition theorem)*

$$\sum_{m=-l}^{l} Y_{lm}(p) Y_{lm}(q) = \frac{2l+1}{4\pi} P_l^0(p \cdot q). \tag{23}$$

Proof Fix the azimuthal angel $\varphi = 0$ and p be the north pole, i.e. $p = (0, 0, 1)$. Now by varying $q = (\sin\theta, 0, \cos\theta)$ for $0 \leq \theta \leq \pi$, we have $Y_{lm}(\theta, \varphi) = 0$ if $m \neq 0$. Then we have

$$\sum_{m=-l}^{l} Y_{lm}(p)Y_{lm}(q) = Y_{l0}(p)Y_{l0}(q) = \frac{2l+1}{4\pi} P_l^0(1)P_l^0(\cos\theta).$$

Note that $P_l^0(p) = 1$. This implies that the sum of product of the spherical harmonics is a function of the inner product between p and q only. From symmetry, if we rotate p back to the original position from the north pole, the same result should hold.

Using the harmonic addition theorem, we simplify the heat kernel in the following theorem.

Theorem 7 *For any $p, q \in S^2$,*

$$K_t^k(p, q) = \sum_{l=0}^{k} \frac{2l+1}{4\pi} e^{-l(l+1)t} P_l^0(p \cdot q). \tag{24}$$

Theorem 7 is used to plot the shape of the heat kernel by fixing p to be the north pole and by varying $\theta = \cos^{-1}(p \cdot q)$ (Figure 3). Similar result is also given in [7]. The maximum of the kernel is obtained at $= 0$. Then the FWHM is solved numerically for θ in

$$\frac{1}{2}\sum_{l=0}^{k} \frac{2l+1}{4\pi} e^{-l(l+1)\sigma} = \sum_{l=0}^{k} \frac{2l+1}{4\pi} e^{-l(l+1)\sigma} P_l^0(\cos\theta).$$

The FWHM is then 2θ. Figure 4 shows the nonlinear relationship between bandwidth t and the corresponding FWHM. When $t = 0.0001$, the corresponding FWHM is 0.2262. This is the FWHM we have used in the subsequent cortical thickness analysis.

5 Unified Surface-based Morphometry via WFS

In this section, we show how the WFS representation is used to quantify cortical surface shape variations in a group of autistic subjects. The WFS provides a unified framework for comparing VBM and the cortical thickness analysis in parallel.

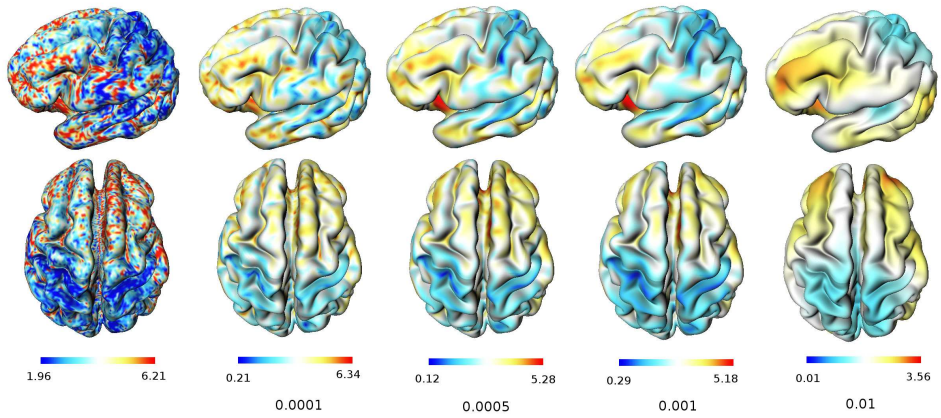


Figure 5: Cortical thickness measures projected onto an average surface for better visualization. As the bandwidth increases from $t = 0.0001$ to $t = 0.01$, the amount of smoothing also increases. The first image shows the cortical thickness obtained from the traditional deformable surface algorithm [11] [13] [36].

5.1 Data Set

$n_1 = 12$ high functioning autistic (HFA) and $n_2 = 12$ normal control (NC) subjects were screened to be right-handed males. The autistic subjects were diagnosed via The Autism Diagnostic Interview - Revised (ADI-R) used by a trained and certified psychologist at the Waisman center at the University of Wisconsin-Madison [18]. Age distributions for HFA and NC are compatible at 15.93 ± 4.71 and 17.08 ± 2.78 respectively.

High resolution anatomical magnetic resonance images (MRI) were obtained using a 3-Tesla GE SIGNA scanner with a quadrature head RF coil. A three-dimensional, spoiled gradient-echo (SPGR) pulse sequence was used to generate T_1-weighted images. Image intensity nonuniformity was corrected using the nonparametric nonuniform intensity normalization method [50] and then the image was spatially normalized into the Montreal neurological institute (MNI) stereotaxic space using a global affine transformation [14]. Afterwards, an automatic tissue-segmentation algorithm based on a supervised artificial neural network classifier was used to classify each voxel as cerebrospinal fluid (CSF), gray matter, or white matter [35]. Subsequently a deformable surface algorithm [36] is used to generate the outer and the inner cortical meshes.

5.2 Stochastic Model

Let $\mathcal{M}_o$ and $\mathcal{M}_i$ be the outer (pial) and inner surfaces of the brain respectively. The unit sphere S^2 is realized as a triangle mesh and deformed to match the outer and inner surfaces in such a way that anatomical homology and the topological connectivity of meshes are preserved [36]. The cortical surfaces can be assumed to be smooth 2-dimensional Riemannian manifolds parameterized by two parameters [19] [33]. Based on the deformable algorithm [36] that establishes the homology between the S^2 mesh and the outer cortical surface, the Cartesian coordinates of the mapping are discretely parameterized by the spherical parametrization (12) as

$$v = (v_1(p), v_2(p), v_2(p)).$$

The inner surface is parameterized similarly as

$$w = (w_1(p), w_2(p), w_3(p)).$$

These discrete coordinate functions are further smoothed by the WFS:

$$v_i(p) = \sum_{l=0}^{k} \sum_{m=-l}^{l} e^{-l(l+1)t} f_{lm}^i Y_{lm}(p). \tag{25}$$

We model v_i stochastically as (22) by assuming f_{lm}^i to follow independent normal distribution $N(\mu_{lm}^i, \sigma_l^2)$ for coordinate i, degree l, and order m. This assumption is equivalent to modeling v_i as the sum of signal plus noise:

$$v_i(p) = \sum_{l=0}^{k} \sum_{m=-l}^{l} e^{-l(l+1)t} \mu_{lm}^i Y_{lm}(p) + \epsilon_i(p),$$

where ϵ_i is a zero mean Guassian random field with a certain isotropic covariance function. A similar stochastic modeling approach has been used in [39] where the canonical expansion of Gaussian random field is used to model the component of a deformation field.

The mean and the variance functions of the surface are given by

$$\begin{aligned} \mathbb{E}v_i(p) &= \sum_{l=0}^{k} \sum_{m=-l}^{l} e^{-l(l+1)t} \mu_{lm}^i Y_{lm}(p), \qquad (26) \\ \mathbb{V}v_i(p) &= \sum_{l=0}^{k} \sum_{m=-l}^{l} e^{-2l(l+1)t} \sigma_l^2 Y_{lm}^2(p) \end{aligned}$$

The total variability of the surface is then measured by

$$\int_{S^2} \mathbb{V} v_i \, d\mu(p) = \sum_{l=0}^{k} \sum_{m=-l}^{l} e^{-2l(l+1)t} \sigma_l^2$$

indicating the increase of smoothing bandwidth decreases the total variability. If

$$v_{ij}(\theta, \varphi) = \sum_{l=0}^{k} \sum_{m=-l}^{l} e^{-l(l+1)t} f_{lm}^{ij} Y_{lm}(\theta, \varphi) \tag{27}$$

is the WFS for the j-th subject ($1 \leq j \leq s$), the unknown parameters μ_{lm}^i and σ_l^2 are estimated as the sample mean and the sample variance:

$$\begin{aligned} \widehat{\mu_{lm}^i} &= \frac{1}{s} \sum_{j=1}^{n} f_{lm}^{ij}, \qquad (28) \\ \widehat{\sigma_l^2} &= \frac{1}{(2l+1)(s-1)} \sum_{m=-l}^{l} \sum_{j=1}^{s} (f_{lm}^{ij} - \mu_{lm}^i)^2. \end{aligned}$$

The inner surface is stochastically modeled similarly as

$$w_i(p) = \sum_{l=0}^{k} \sum_{m=-l}^{l} e^{-l(l+1)t} g_{lm}^i Y_{lm}(p). \tag{29}$$

5.3 Surface normalization

Previously cortical surface normalization is performed by minimizing an objective function that measures the global fit of two surfaces while maximizing the smoothness of the deformation in such a way that the gyral patterns are matched smoothly [11] [47] [52]. In the WFS representation, the surface normalization is straightforward and does not require any sort of optimizations explicitly.

Given surfaces v_{i1} and v_{i2} as in (27), the displacement field d_i that minimizes the integral of the squared errors of warping v_{i1} to v_{i2} is simply given by the following theorem.

Theorem 8

$$v_{i2} - v_{i1} = \arg \min_{d_i \in \mathcal{H}_k} \int_{S^2} [d_i(v_{i1}) - v_{i2}]^2 \, d\mu(p).$$

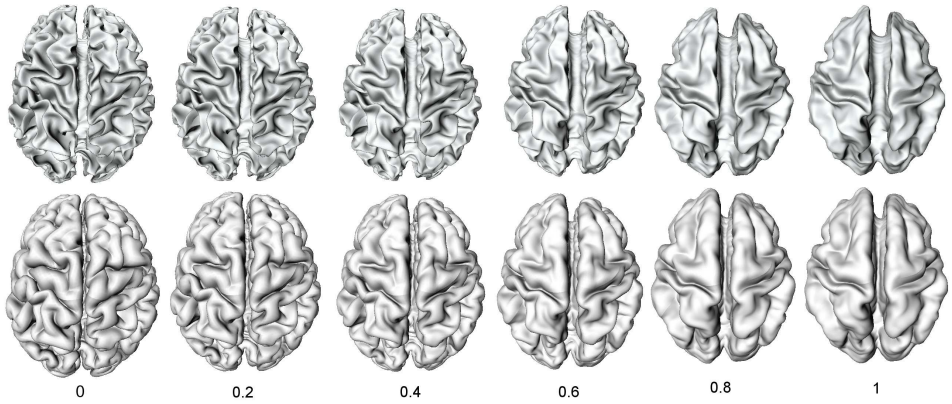


Figure 6: Multi-scale representation of surface registration toward the average template. The top is the inner surface while the bottom is the outer surface. $\alpha = 0$ is the surface of one particular subject while $\alpha = 1$ is the average surface of 24 subject. The amount of registration serves as a multi-scale representation where the optimal scale for differentiating two groups should be searched for.

Theorem 8 shows that the optimal displacement in the least squares sense is obtained by simply taking the difference between two WFS representations. Unlike other surface registration methods used in warping surfaces between subjects [11] [47] [52], it is not necessary to consider an additional cost function that guarantees the smoothness of the displacement field since the displacement field $v_{i2} - v_{i1}$ is already a linear combination of smooth basis functions. Based on this idea, we normalize WFS surfaces.

Let $\bar{v}_i$ be the mean surface obtained by replacing μ^i_{lm} in (26) with the sample mean (28). Figure 6 shows the mean surface for 24 subjects used in the study. The mean surface serves as a template for a statistical analysis later. For subject j, the displacement from surface v_{ij} to the template is

$$\Delta v_{ij} = \bar{v}_i - v_{ij}.$$

Consider surface

$$\mathbf{v}_{ij}(\alpha) = v_{ij} + \alpha \Delta v_i = (1 - \alpha) v_{ij} + \alpha \bar{v}_i, \tag{30}$$

which is the trajectory of the deformation from v_{ij} to the template $\bar{v}_i$ parameterized by $\alpha \in [0, 1]$. When $\alpha = 0$, $\mathbf{v}_{ij}(\alpha)$ is the j-th subject surface while

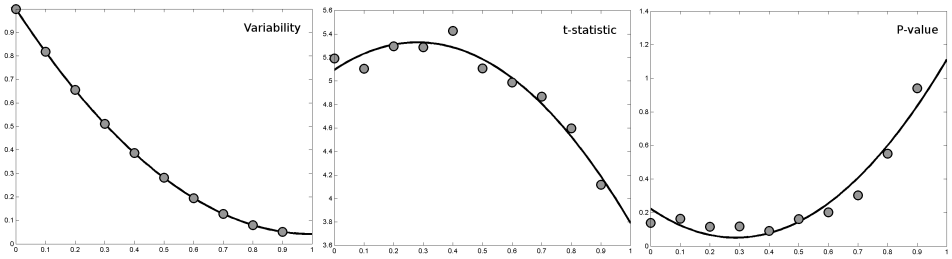


Figure 7: Left: plot of image registration variability $c(\alpha)$. Middle: plot of the maximum of T statistic at each scale. Right: plot of corrected P-value corresponding to the maximum of T statistic. At $\alpha = 0.4$, the minimum P-value of less than 0.1 is obtained so we choose $\alpha = 0.4$ to be the optimal scale that separates the two groups. The lines are the best fitting quadratic curve in the least squares sense.

when $\alpha = 1$ it is the template surface. The parameter α controls the amount of registration from the coarse-to-fine scale toward the template. Figure 6 shows $\mathbf{v}_i(\alpha)$ at 11 different scales between 0 and 1 with 0.1 increment for a single subject. The larger the value of α, the smaller the image registration variability across the subjects with respect to the template. This is shown from the total variability computed at each scale α:

$$\int_{S^2} \mathbb{V}(\mathbf{v}_{ij})\, d\mu(p) = c(\alpha) \sum_{l=0}^{k} \sum_{m=-l}^{l} e^{-2l(l+1)t} \sigma_l^2,$$

where

$$c(\alpha) = \frac{n-1}{n^2}\alpha^2 + \big(1 - \frac{n-1}{n}\alpha\big)^2$$

is decreasing over $0 \leq \alpha \leq 1$ (Figure 7).

5.4 Optimal Voxel-based Morphometry

Based on the multi-scale representation of surface registration, the optimal amount of registration needed in VBM is determined. The gray matter density is constructed using the 3D Euclidian distance map of the surfaces at each scale. For the outer surface $\mathcal{M}_o(\alpha)$ at scale α, the distance map at each voxel x is defined as

$$\texttt{dist}_o(x) = \min_{y \in \mathcal{M}_o(\alpha)} \|x - y\|,$$

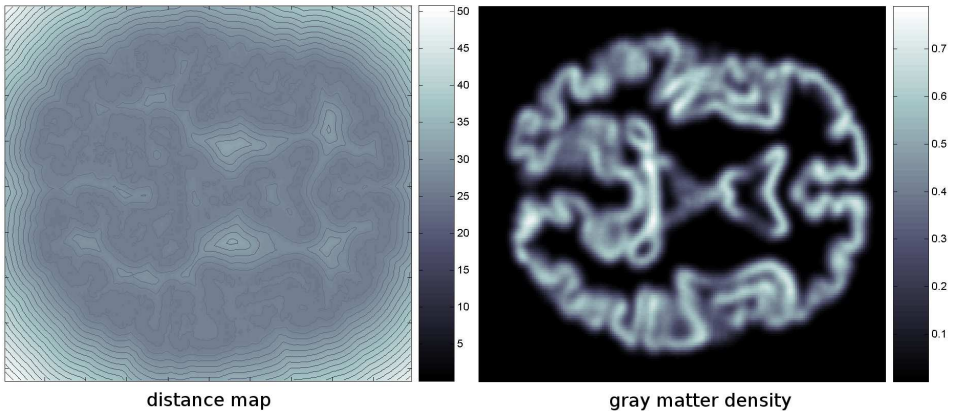


Figure 8: Left: contour plot of the average distance map in mm at scale $\alpha = 0$. Right: Gaussian kernel smoothing of the gray matter density map with 10mm FWHM.

where $\|\cdot\|$ is the Euclidian norm. The minimum is found using the nearest neighbor search algorithm on an optimized k-D tree [25]. Similarly we denote the distance map for the inner surface $\mathcal{M}_i(\alpha)$ as $\texttt{dist}_i(x)$. Then the average distance map is defined as

$$\texttt{dist}(x) = \frac{\texttt{dist}_o(x) + \texttt{dist}_i(x)}{2}.$$

The average distance map for a subject is shown in Figure 8. The minimum of the average distance is always obtained in the middle of the outer and the inner surfaces, where the probability of a voxel belong to the gray matter class should be the highest. Then we define the gray matter density as

$$\texttt{density}(x) = \exp\Big[-\frac{\texttt{dist}^2(x)}{2\rho^2}\Big], \tag{31}$$

where parameter ρ^2 controls the spread of density. In this paper, we used $\rho^2 = 3$. The gray matter density is always between 0 and 1 and it obtains its maximum in the interior of the gray matter region, where the average distance map obtains the minimum. The density map is further convoluted with the 3D Gaussian kernel K with 10mm FWHM to increase the smoothness and normality of data [2] [10] (Figure 8). The smoothed density map $K * \texttt{density}(x)$ is stochastically modeled as a Gaussian random field.

At each scale, we construct the smoothed density map for each subject. Then the two sample t-test statistic T with the equal variance assumption is computed on the convoluted gray matter density maps at each voxel [10] . The resulting T random field $T(x)$ is distributed as a student t distribution with $\nu = n_1 + n_2 - 2$ degrees of freedom at each voxel x. Based on the random field theory [56] [55] [54], the test statistic, which accounts for the multiple comparison correction, is the maxima of T field over the gray matter $\mathcal{M}_g$. The corresponding *corrected P-value* is computed using the following formula:

$$\begin{aligned} & P\Big(\sup_{x\in\mathcal{M}_g} T(x) > h\Big) \\ \approx \; & \frac{\mathrm{Vol}(\mathcal{M}_g)}{\mathrm{FWHM}^3}\frac{(4\ln 2)^{3/2}}{(2\pi)^2}\Big(\frac{\nu-1}{\nu}h^2-1\Big)\Big(1+\frac{h^2}{\nu}\Big)^{-\frac{\nu-1}{2}}, \end{aligned}$$

where $\mathrm{Vol}(\mathcal{M}_g) = 2.13 \times 10^5 \mathrm{mm}^3$ is the volume of the gray matter of the template. The gray matter volume is estimated by computing the volume bounded by the outer and inner meshes [13]. Restricting the search region from the whole brain volume to the gray matter boosts the signal detection power. The optimal image registration scale is determined to be the one that provides the maximal discrepancy between the groups. Hence, the minimum corrected P-value can be chosen as a criteria for determining the optimal scale. The maximum T statistic value and its corresponding corrected P-value at each scale are plotted in Figure 7 showing that the optimal scale is obtained when $\alpha = 0.4$. At this scale, the maximum T-stat. is 5.43 while the minimum T-stat. is -5.04. The random field theory based thresholding of $h = \pm 5.35$ gives the corrected P-value of 0.1. Figure 9 shows the optimally constructed T-stat. map thresholded at ± 4.0 and interpolated into the nearest point in the cortical surface showing increased gray matter density in the localized areas of the autistic subjects.

5.5 Cortical Thickness Analysis

The previously available approaches for computing the cortical thickness in discrete triangle meshes produce noisy thickness measures [11] [23] [36]. So it is necessary to smooth the thickness measurements along the cortex via PDE based smoothing techniques [1] [8] [13]. On the other hand, the WFS provides smooth functional representation of the outer and inner surfaces so that the distance measures between the surfaces should be already smooth.Hence, the WFS avoids this additional step of thickness smoothing

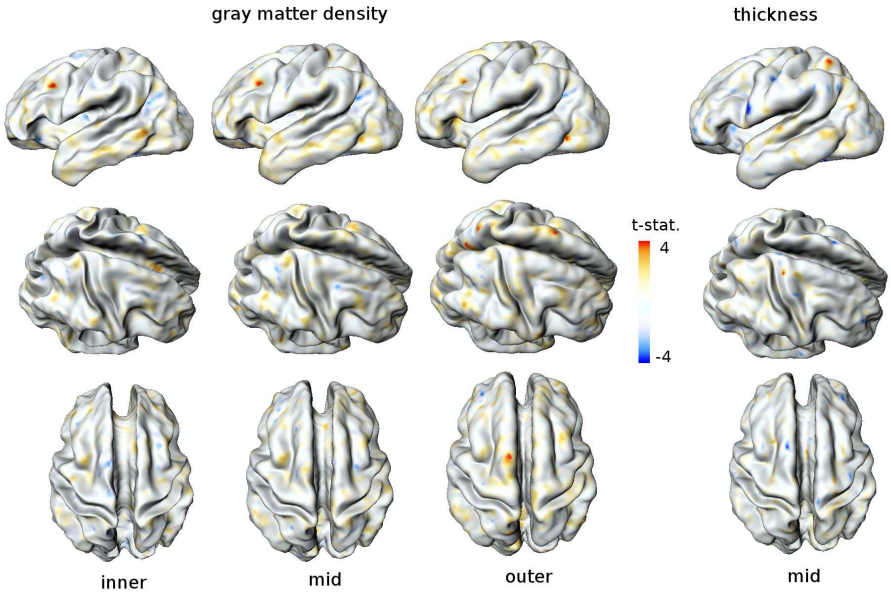


Figure 9: T statistic map for gray matter density and cortical thickness thresholded at between -4 and 4 for the comparison purpose. In the gray matter density results, signals are mainly detected in either the inner surface or the outer surface but not in the middle surface. This validates the conjecture in VBM that the signal will be mainly detected around the tissue boundary. The gray matter density results and the thickness results do not overlap reconfirming that the thickness and gray matter density are not positive correlated measurements. Our study directly demonstrates that the VBM and cortical thickness analysis SPMs will not overlap.

done in most of thickness analysis literatures [13] [11] . It is not necessary to perform data smoothing in the WFS formulation.

Using Theorem 8, we establish the homology between the outer and the inner surfaces in the least squares fashion. We will call this homology as the *WFS-correspondence*. For the outer surface (25) and the inner surface (29), the cortical thickness is defined to be the Euclidean distance between the WFS-correspondence:

$$\texttt{thick}(p) = \Big[\sum_{l=0}^{k}\sum_{m=-l}^{l} e^{-2l(l+1)t}(g_{lm}^{i} - f_{lm}^{i})^2\Big]^{1/2}.$$

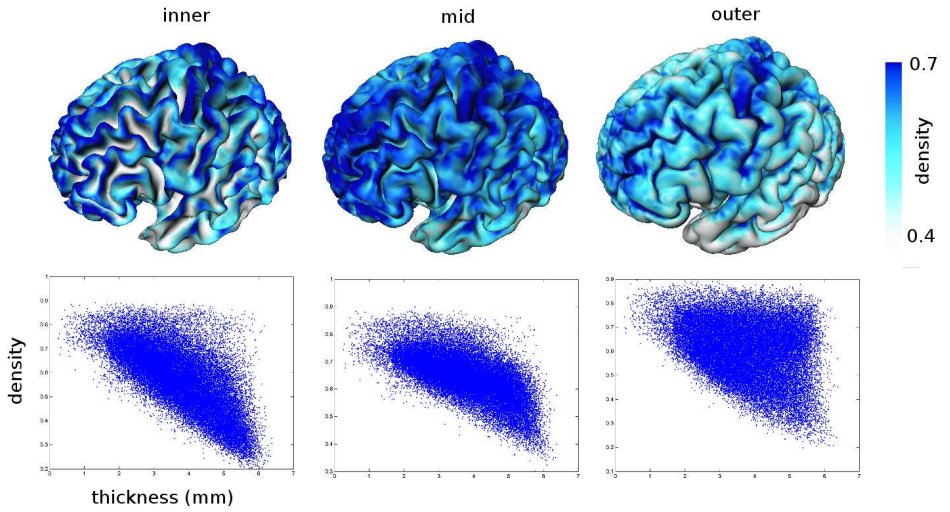


Figure 10: Top: gray matter density projected onto inner, mid and outer surfaces. On the inner surface, the deep sulcal regions show the low density while the gyral ridges show high density. On the outer surface, this is opposite. The deep sulcal regions show high density while the gyral ridges show lower density. The middle surface shows high density. Bottom: Scatter plot of gray matter density over thickness. They show negative correlations.

A similar approach has been proposed for measuring the closeness between two surfaces [27]; however, this is the first study using the spherical harmonics in defining the cortical thickness. Figure 5 shows the comparison of cortical thickness computed from the traditional deformable surface algorithm [36] and the WFS-correspondence. The cortical thickness obtained from the traditional approach introduces a lot of triangle mesh noise into its estimation while the WFS-correspondance approach dose not. The spatial smoothness of the thickness is controlled by the bandwidth t.

For the group comparision between the autistic and the normal control groups, two sample t-test is performed. The corresponding corrected P-value is computed using the following formula:

$$\begin{aligned} & P\Big(\sup_{p\in S^2} T(p) > h\Big) \\ \approx \quad & \frac{1}{\mathrm{FWHM}^2}\frac{(4\ln 2)}{(2\pi)^{1/2}}\frac{\Gamma(\frac{\nu+1}{2})}{(\frac{\nu}{2})^{1/2}\Gamma(\frac{\nu}{2})}h\Big(1+\frac{h^2}{\nu}\Big)^{-\frac{\nu-1}{2}}, \end{aligned}$$

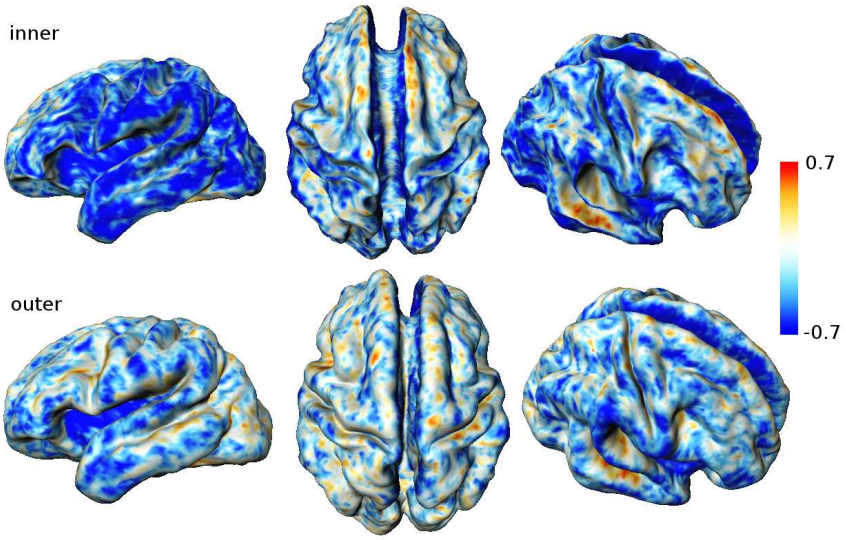
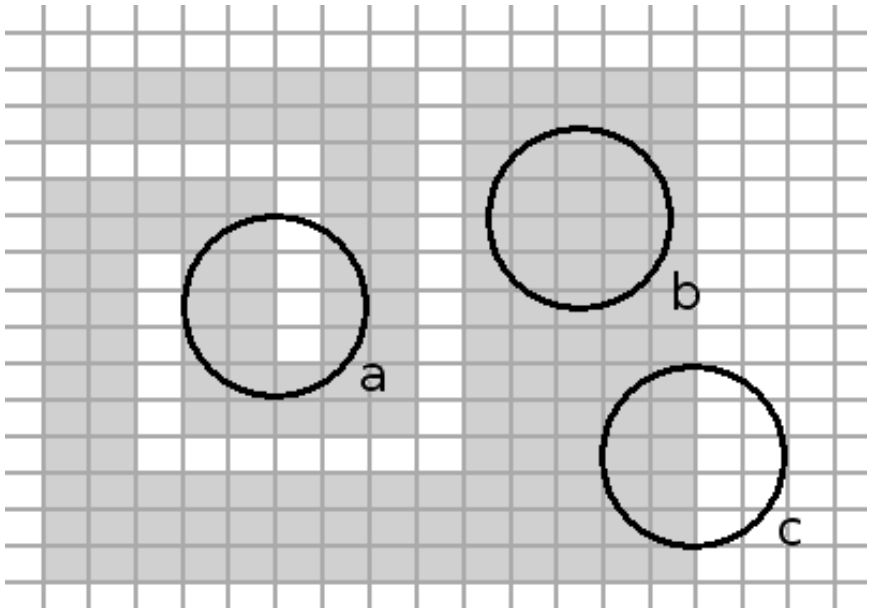


Figure 11: Correlation of thickness and gray matter density for 24 subjects mapped on both the inner and the outer surfaces. Most of both inner and outer surfaces show negative correlation. Thicker cortical regions are less convoluted so the gray matter density tend to be lower.

where FWHM is from the heat kernel used in the WFS. We used FWHM = 0.2262 corresponding to bandwidth $t = 0.0001$. Then for the threshold $h = \pm 4.5$, we obtain the corrected P-value of 0.1. The minimum T-stat. is -4.73 while the maximum T-stat. is 4.83. Figure 9 shows the T-stat. map thresholded at ± 4.

5.6 Comparing Cortical Thickness and Gray Matter Density

Most morphometric studies [2] [5] [11] [13] [23] [29] perform VBM and cortical thickness analysis separately and it is not clear if two anatomical indices measure the same anatomical characteristic. Although both the gray matter density and the cortical thickness are presumed to measure the amount of gray matter, it is unclear if these two measures are positively correlated. Comparing the SPM of density and thickness in Figure 9, no statistically significant regions overlap. Since both metrics have been assumed to be the indicators of the amount of gray matter, the result is paradoxical. So we have correlated these two metrics within a subject (Figure 10) and across



[t]

Figure 12: Simple 2D schematic showing the negative correlation between thickness and gray matter density. Gray colored pixels are the gray matter. The black circles are the contour of heat kernel. There are less gray matter pixels in region (c) than region (a) although the thickness in region (c) is thicker than that of region (a). The gray matter density in the middle of the gray matter (b) is close to 1 for all subject indicating very small between-subject and between-group variability. Because of the small between-group variability, VBM does not usually detect signal in the middle of the gray matter. Most of the significant signal detected in VBM is near the tissue boundary where the between-group variability is high.

subjects (Figure 11) to further investigate the relationship. Surprisingly the scatter plot in Figure 10 shows negative correlation within a subject. By assigning the density value of a voxel that contains a vertex of a cortical mesh to the vertex, we can project the gray matter density onto inner, middle and outer surfaces. The middle surface is obtained by averaging the inner and the outer surfaces in the WFS-correspondance. Surfaces show different patterns of a negatively correlated scatter plots. Figure 11 shows complex pattern of nonuniformity of density. On the outer surface, deep sulci have higher density compared to gyri while on the inner surface, the pattern is opposite. The middle surface shows higher density compared to the outer and the inner surfaces as expected. These complex patterns of the nonuniformity of density is due to the folding pattern of the cortex. Since the sulci on the outer surface and the gyri on the inner surface are highly folded, these regions should have more gray matter within the sphere of fixed radius as illustrated in Figure 12. On the other hand, thin cortical regions will fold more than thick cortical regions. This inverse geometric relation is causing the negative correlation between density and thickness and, in turn, the resulting SPM differ in the regions of statistically significant difference. We further computed the correlation between two measures across 24 subjects. Figure 10 shows a similar result showing negative correlation across subjects in most regions of cortex. Based on this primary result, it is clear that gray matter density and cortical thickness should be analyzed together in a multivariate fashion rather than analyzing separately.

6 Conclusions and Discussions

In this paper, we presented a unified theoretical framework for the WFS and detailed numerical implementation issues. The WFS is used as a smooth global parametrization of the cortex. It is a very flexible functional estimation technique for scalar and vector data projected onto a unit sphere. The WFS can be also viewed as a cortical data smoothing technique. The WFS is shown to be a solution of the Cauchy problem in PDE and for a specific weights, it becomes diffusion smoothing [13]. As an application of this novel approach, we used the WFS as a tool for comparing the VBM and the cortical thickness analysis. Using the WFS representation as the ground truth, cortical thickness and gray matter density are constructed, and morphometric analysis on these indices are performed.

In the VBM, the problem of determining optimal amount of image registration has been addressed. It was shown that the opimal scale is obtained

somewhere between global affine registration and the WFS-correspondance based nonlinear warping. This multi-scale VBM incorporates the convoluted nature of the gray matter using the WFS more accurately than the previous 3D-based VBM. The explicit mathematical representation of the WFS-based surface-to-surface registration enabled us to construct the trajectory of the deformation field. This trajectory is used as a parameter for controlling the amount of image registration in a multi-scale fashion. Then the optimal VBM is chosen that gives the maximal discrimination between the two clinical groups. In the cortical thickness analysis, the thickness is established using the same WFS- correspondence. Afterwards, the SPMs of the VBM and the thickness are compared to show the statistically significant regions do not overlap. This surprising result is caused by the negative correlation between density and thickness. Increased folding increases the gray matter density while decreasing thickness. This should serve as a spring board for investigating further on comparing the VBM and the cortical thickness analysis.

Acknowledgment

The authors wish to thank Steven Robbins of the Montreal Neurological Institute, John Ashburner of the Wellcome Department of Imaging Neuroscience, Martin A. Syner of the Department of Computer Science, University of North Carolina at Chapel Hill, and Shubing Wang of the Department of Statistics, University of Wisconsin-Madison for valuable discussions and comments.

References

[1] A. Andrade, F Kherif, J. Mangin, K.J. Worsley, A. Paradis, O. Simon, S. Dehaene, D. Le Bihan, and J-B. Poline. Detection of fmri activation using cortical surface mapping. *Human Brain Mapping*, 12:79–93, 2001.

[2] J. Ashburner and K. Friston. Voxel-based morphometry - the methods. *NeuroImage*, 11:805–821, 2000.

[3] J. Ashburner and K. Friston. Why voxel-based morphometry should be used. *NeuroImage*, 14:1238–1243, 2001.

[4] J Ashburner, C. Good, and K.J. Friston. Tensor based morphometry. *NeuroImage*, 11S:465, 2000.

[5] F.L. Bookstein. Voxel-based morphometry should not be used with imperfectly registered images. *NeuroImage*, 14:1454–1462, 2001.

[6] C. Brechbuhler, G. Gerig, and O. Kubler. Parametrization of closed surfaces for 3D shape description. *Computer Vision and Image Understanding*, 61:154–170, 1995.

[7] T. Bulow. Spherical diffusion for 3d surface smoothing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:1650–1654, 2004.

[8] A. Cachia, J.-F. Mangin, Riviére D., D. Papadopoulos-Orfanos, F. Kherif, I. Bloch, and J. Régis. A generic framework for parcellation of the cortical surface into gyri using geodesic voronoï diagrams. *Image Analysis*, 7:403–416, 2003.

[9] M.K. Chung. Heat kernel smoothing on unit sphere. In *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, 2006.

[10] M.K. Chung, K.M. Dalton, A.L. Alexander, and R.J. Davidson. Less white matter concentration in autism: 2D voxel-based morphometry. *NeuroImage*, 23:242–251, 2004.

[11] M.K. Chung, S. Robbins, Davidson R.J. Alexander A.L. Dalton, K.M., and A.C. Evans. Cortical thickness analysis in autism with heat kernel smoothing. *NeuroImage*, 25:1256–1265, 2005.

[12] M.K. Chung, S. Robbins, K.M. Dalton, S. Wang, A.C. Evans, and R.J. Davidson. Tensor-based cortical morphometry via weighted spherical harmonic representation. In *IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA)*, 2006.

[13] M.K. Chung, K.J. Worsley, S. Robbins, T. Paus, Taylor, J.N. J., Giedd, J.L. Rapoport, and A.C. Evans. Deformation-based surface morphometry applied to gray matter deformation. *NeuroImage*, 18:198–213, 2003.

[14] D.L. Collins, P. Neelin, T.M. Peters, and A.C. Evans. Automatic 3d intersubject registration of mr volumetric data in standardized talairach space. *J. Comput. Assisted Tomogr.*, 18:192–205, 1994.

[15] R. Courant and D. Hilbert. *Methods of Mathematical Physics, Volume 1.* Interscience, New York, 1953.

[16] W.R. Crum, L.D. Griffin, D.L.G. Hill, and D.J. Hawkes. Zen and the art of medical image registration: correspondence, homology, and quality. *NeuroImage*, 20:1425–1437, 2003.

[17] A.M. Dale and B. Fischl. Cortical surface-based analysis i. segmentation and surface reconstruction. *NeuroImage*, 9:179–194, 1999.

[18] K.M. Dalton, B.M. Nacewicz, T. Johnstone, H.S. Schaefer, M.A. Gernsbacher, H.H. Goldsmith, A.L. Alexander, and R.J. Davidson. Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience*, 8:519–526, 2005.

[19] C. Davatzikos and R.N. Bryan. Using a deformable surface model to obtain a shape representation of the cortex. *Proceedings of the IEEE International Conference on Computer Vision*, 9:2122–2127, 1995.

[20] C. Davatzikos, A. Genc, D. Xu, and S.M. Resnick. Voxel-based morphometry using the ravens maps: Methods and validation using simulated longitudinal atrophy. *NeuroImage*, 14:1361–1369, 2001.

[21] C. Davatzikos, M. Vaillant, S.M. Resnick, J.L. Prince, S. Letovsky, and N Bryan. A computerized approach for morphological analysis of the corpus callosum. *Journal of Computer Assisted Tomography*, 20:88–97, 1996.

[22] J. Fan and I. Gijbels. *Local Polynomial Modelling and Its Applications*. Chapman & Hall/CRC, 1996.

[23] B. Fischl and A.M. Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *PNAS*, 97:11050–11055, 2000.

[24] B. Fischl, M.I. Sereno, R. Tootell, and A.M. Dale. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapping*, 8:272–284, 1999.

[25] J.H. Friedman, J.L. Bentley, and R.A. Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM transactions on mathematics software*, 3:209–226, 1997.

[26] K.J. Friston. A short history of statistical parametric mapping in functional neuroimaging. Technical Report Technical report, Wellcome Department of Imaging Neuroscience, ION, UCL., London, UK., 2002.

[27] G. Gerig, M. Styner, D. Jones, D. Weinberger, and J. Lieberman. Shape analysis of brain ventricles using spharm. In *MMBIA*, pages 171–178, 2001.

[28] G. Gerig, M. Styner, and G. Szekely. Statistical shape models for segmentation and structural analysis. I:467–473, 2004.

[29] C.D. Good, I.S. Johnsrude, J. Ashburner, R.N.A. Henson, K.J. Friston, and R.S.J. Frackowiak. A voxel-based morphometric study of ageing in 465 normal adult human brains. *NeuroImage*, 14:21–36, 2001.

[30] H. Groemer. *Geometric Applications of Fourier Series and Spherical Harmonics,*. Cambridge University Press, 1996.

[31] X. Gu, Y.L. Wang, T.F. Chan, T.M. Thompson, and Yau. S.T. Genus zero surface conformal mapping and its application to brain surface mapping. *IEEE Transactions on Medical Imaging*, 23:1–10, 2004.

[32] S.C. Johnson, L.C. Baxter, L. Susskind-Wilder, D.J. Connor, M.N. Sabbagh, and R.J. Caselli. Hippocampal adaptation to face repetition in healthy elderly and mild cognitive impairment. *Neuropsychologia*, 42:980–989, 2004.

[33] S.C. Joshi, J. Wang, M.I. Miller, D.C. Van Essen, and U. Grenander. On the differential geometry of the cortical surface. *Vision Geometry IV*, pages 304–311, 1995.

[34] A. Kelemen, G. Szekely, and G. Gerig. Elastic model-based segmentation of 3-d neuroradiological data sets. *IEEE Transactions on Medical Imaging*, 18:828–839, 1999.

[35] K. Kollakian. Performance analysis of automatic techniques for tissue classification in magnetic resonance images of the human brain. Technical Report Master's thesis, Concordia University, Montreal, Quebec, Canada, 1996.

[36] J.D. MacDonald, N. Kabani, D. Avis, and A.C. Evans. Automated 3-d extraction of inner and outer surfaces of cerebral cortex from mri. *NeuroImage*, 12:340–356, 2000.

[37] A.B. McMillan, B.P. Hermann, S.C. Johnson, R.R. Hansen, M. Seidenberg, and M.E. Meyerand. Voxel-based morphometry of unilateral temporal lobe epilepsy reveals abnormalities in cerebral white matter. *NeuroImage*, 23:167–174, 2004.

[38] S. Mehta, T.J. Grabowski, Y. Trivedi, and H. Damasio. Evaluation of voxel-based morphometry for focal lesion detection in individuals. *NeuroImage*, 20:1438–1454., 2004.

[39] M.I. Miller, A. Banerjee, G.E. Christensen, S.C. Joshi, N. Khaneja, U. Grenander, and L. Matejic. Statistical methods in computational anatomy. *Statistical Methods in Medical Research*, 6:267–299, 1997.

[40] M.I. Miller, A.B. Massie, J.T. Ratnanather, K.N. Botteron, and J.G. Csernansky. Bayesian construction of geometrically based cortical thickness metrics. *NeuroImage*, 12:676–687, 2000.

[41] H.-G. Muller. Functional modeling and classification of longitudinal data. *Scandinavian Journal of Statistics*, 32:223–240, 2005.

[42] T. Paus, A. Zijdenbos, K.J. Worsley, D.L. Collins, J. Blumenthal, J.N. Giedd, J.L. Rapoport, and A.C. Evans. Structural maturation of neural pathways in children and adolescents: In vivo study. *Science*, 283:1908–1911, 1999.

[43] D.A. Pizzagalli, T.R. Oakes, A.S. Fox, M.K. Chung, C.L. Larson, H.C. Abercrombie, S.M. Schaefer, R.M. Benca, and R.J. Davidson. Functional but not structural subgenual prefrontal cortex abnormalities in melancholia. *Molecular Psychiatry*, 9:393–405, 2004.

[44] A. Qiu, D. Bitouk, and M.I. Miller. Smooth functional and structural maps on the neocortex via orthonormal bases of the laplace-beltrami operator. *IEEE Transactions on Medical Imaging*, 2006. in press.

[45] M. Quicken, C Brechbuhler, J. Hug, H. Blattmann, and G. Szekely. Parameterization of closed surfaces for parametric surface description. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 354–360, 2000.

[46] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer, 1997.

[47] S.M. Robbins. Anatomical standardization of the human brain in euclidean 3-space and on the cortical 2-manifold. Technical Report PhD thesis, School of Computer Science, McGill University, Montreal, Quebec, Canada, 2003.

[48] L. Shen and M.K. Chung. Large-scale modeling of parametric surfaces using spherical harmonics. In *Third International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)*.

[49] L. Shen, J. Ford, F. Makedon, and A. Saykin. surface-based approach for classification of 3d neuroanatomical structures. *Intelligent Data Analysis*, 8, 2004.

[50] J.G. Sled, A.P. Zijdenbos, and A.C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE Transactions on Medical Imaging*, 17:87–97, 1988.

[51] P.M. Thompson, Hayashi K.M., de Zubicaray G., Janke A.L., Rose S.E., Semple J., Herman D., Hong M.S., Dittmer S.S., Doddrell D.M., and Toga A.W. Dynamics of gray matter loss in alzheimer's disease. *J. Neurosci.*, 23:994–1005, 2003.

[52] P.M. Thompson and A.W. Toga. A surface-based technique for warping 3-dimensional images of the brain. *IEEE Transactions on Medical Imaging*, 15, 1996.

[53] G. Wahba. *Spline models for observational data*. SIAM, 1990.

[54] K.J. Worsley. Local maxima and the expected euler characteristic of excursion sets of $\chi 2$, f and t fields. *Advances in Applied Probability.*, 26:13–42, 1994.

[55] K.J. Worsley, S. Marrett, P. Neelin, A.C. Vandal, K.J. Friston, and A.C. Evans. A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, 4:58–73, 1996.

[56] K.J. Worsley, J.E. Taylor, F. Tomaiuolo, and J. Lerch. Unified univariate and multivariate random field theory. *NeuroImage*, 12:in press, 2005.

[57] I. C. Wright, P. K. McGuire, J.-B. Poline, J. M. Travere, R. M. Murray, C. D. Frith, R. S. J. Frackowiak, and K. J. Friston. A voxel-based method for the statistical analysis of gray and white matter density applied to schizophrenia. *NeuroImage*, 2:244–252, 1995.